The Design and Development of a Potent and Selective Novel Diprolyl Derivative that Binds to the N-domain of Angiotensin-I Converting Enzyme

Journal:	*Journal of Medicinal Chemistry*
Manuscript ID	Draft
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Fienberg, Stephen; University of Cape Town, Chemistry Cozier, Gyles; University of Bath, Acharya, K. R. ; University of Bath, Department of Biology and Biochemistry Chibale, Kelly; University of Cape Town, Department of Chemistry Sturrock, Edward; University of Cape Town, Integrative Biomedical Sciences

The Design and Development of a Potent and Selective Novel Diprolyl Derivative that Binds to the N-domain of Angiotensin-I Converting Enzyme

Stephen Fienberg[a], Gyles E. Cozier[b], K. Ravi Acharya[b], Kelly Chibale[a,c,e]* and Edward D. Sturrock[c,d]*

[a] Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[b] Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK

[c] Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Rondebosch 7701, South Africa

[d] Department of Integrative Biomedical Sciences, University of Cape Town, Observatory 7925, South Africa

[e] South African Medical Research Council Drug Discovery and Development Research Unit,, University of Cape Town, Rondebosch 7701, South Africa

*To whom correspondence should be addressed.

Edward Sturrock, phone, +27 21 4066312; fax, +27 21 4066479; e-mail, Edward.Sturrock@uct.ac.za;

Kelly Chibale, Phone, + 27 21 6502553; fax, 27 21 6504521; e-mail, Kelly.Chibale@uct.ac.za

ABSTRACT

Angiotensin-I converting enzyme (ACE) is a dual-domain zinc metalloprotease best known for its critical role in blood pressure regulation. The two domains of ACE (N- and C-) both hydrolyse angiotensin-I and bradykinin, but show different substrate and inhibitor specificities with the antifibrotic tetrapeptide Ac-SDKP being highly selective for the N-domain. Until recently, all clinical ACE inhibitor(s) (ACEi) have been shown to be non-selective towards either domain. The entries in the GVK (Gunupati Venkata Krishna) Biosciences database pertaining to ACE were inspected for possible N-domain selective binding patterns. From this set, a series of diprolyl compounds was selected and modelled using docking simulations. The series was expanded upon to target key interactions involving residues known to impart N-domain selectivity. A selection of seven diprolyl compounds were synthesised and tested for N-domain selective ACE inhibition. One compound with an aspartic acid in the P_2 position (compound **16**) displayed potent inhibition (K_i = 12 nM), and was 84-fold more selective towards the N-domain. A high-resolution crystal structure of compound **16** in complex with the N-domain revealed a molecular basis for the observed selectivity.

INTRODUCTION

Angiotensin-I converting enzyme (ACE; EC 3.4.15.1) is a zinc dipeptidyl carboxypeptidase that cleaves the inactive decapeptide angiotensin-I (Ang-I) into the vasopressor angiotensin-II (Ang-II), and degrades the vasodilator bradykinin (BK). The importance of these activities is emphasised by the extensive use of ACE inhibitor(s) (ACEi) for the treatment of hypertension and cardiovascular disease. The ACEi captopril[1] and lisinopril[2] were designed in the 1970s and later extended by others. Currently, more than 17 ACEi have been approved for clinical use.

Captopril is a peptidomimetic compound designed to mimic the sequence of peptides isolated from *Bothrops jaracara* venom, which were known to inhibit the ACE catalysed dipeptidyl cleavage of Ang-I. Since the elucidation of the RAAS, many more peptide substrates of ACE have been identified and have been shown to play a role in many more signalling pathways.[3]

At the time of the development of captopril, ACE was assumed to share mechanistic similarities with carboxypeptidase A, a zinc metalloprotease of which the structure was known.[4] Following the cloning of somatic ACE, it was shown to consist of two homologous catalytic domains known as the N- and C-domains. Apart from structural differences, the two domains have been shown to possess different kinetic profiles when hydrolysing their peptide substrate.

Subsequent studies have shown the C-domain to be responsible for the majority of Ang-I hydrolysis while the other major substrate bradykinin has been shown to be hydrolysed at a similar rate by both the N- and C-domain.[5] It is currently believed that the two domains of ACE act independently of each other with the total activity of somatic ACE (sACE) equalling the sum of activities for the individual domains.

The tetrapeptide AcSDKP was discovered around the same time that the two domains of ACE were elucidated.[6] The presence of AcSDKP has been shown to reduce the deposition of collagen in haematopoetic stem cells of both cardiac and pulmonary tissues.[7] Excess collagen deposition in these tissues has been identified as a mechanism for the pathogenesis of fibrosis. AcSDKP has been shown to be exclusively hydrolysed by the N-domain of ACE.[8] N-domain selective ACE inhibition has been identified as a possible treatment for both cardiac and pulmonary fibrosis without the side-effects of non-selective inhibition of both domains of ACE.

Individual N- and C-domain ACE constructs have been crystallised with a variety of different inhibitors co-crystallised with both catalytic domains. These crystal structures have clearly elucidated the core interactions responsible for potent ACE inhibition.[9,10]

The N- and C-domains share 65% sequence homology with each other.[11] Many of the key interactions are shared between the two domains with most inhibitors showing similar K_i values. Both substrates and most ACEi bind to the ACE active site via a chelation interaction with the catalytic zinc metal, a H-bond between the catalytic Glutamic acid and the amide nitrogen of the zinc binding amide and a polar salt-bridge interaction between the charged terminal carboxylic acid and three conserved residues in the S_2' subsite (Gln259, Lys489 and Tyr498 in the N-domain, conserved as Gln281, Lys511 and Tyr520 in the C-domain). The P_1' backbone carbonyl interacts with the conserved His331 and His491 in the N-domain (His353 and His513 in the C-domain). A conserved tyrosine (Tyr501 and Tyr523 in the N- and C-domains respectively) can interact with either the P_1'

peptide carbonyl, or the ZBG. Lipophilic interactions between non-polar groups and the non-polar subsites indicated in figure 1 add to the potency of the inhibitor. These are mostly conserved between the two domains. However, the C-domain S_2' is more lipophilic and favours P_2' bicyclic groups, largely due to the replacement of Ser257 and Thr358 in the N-domain with Val379 and Val380 in the C-domain. The S_1 subsite is largely similar except for Thr496 in the N-domain being replaced with Val518 in the C-domain, making this subsite more hydrophobic in the C-domain.

There are various conserved residues in the S_2' subsite, and a conserved interaction between the backbone S_2' carbonyl with the backbone nitrogen of Ala334 in the N-domain (Ala356 in the C-domain). However, there are two key differences between the two binding sites with the N-domain to C-domain substitutions of Arg381 to Glu403 and Tyr369 to Phe391. Both these substitutions could influence ligand binding. The substitution of a positive charge for a negative one, and a hydrophilic residue for a hydrophobic one provides an ideal set of interactions to be targeted for the design of a domain selective ACEi.

Both of these substitutions were shown to impart N-domain selectivity in the case of the phosphinic tetrapeptide RXP407, the first example of an N-domain selective ACEi.[12] RXP407 contains all the core ACEi moieties with the addition of an aspartic acid in the P_2 position. RXP407 was shown to inhibit the N-domain with a 2500-fold K_i selectivity over the C-domain. This selectivity was explained via a P_2 salt bridge with Arg381 and a H-bond partner Tyr369 with the N-domain. Substituting the positively charged Arg381 and the H-bond donor Tyr369 with the negatively-charged Glu403 and the hydrophobic Phe391 in the C-domain could explain the dramatic disruption of binding observed in the C-domain.[13]

While the N-domain selective ACE inhibition observed for RXP407 proved to be an excellent proof-of concept, the compound is too large and peptidic to be considered drug-like. Despite its shortcomings, RXP407 provides a template upon which to base the design of a drug-like N-domain selective ACEi. Designing an ACEi specific to just one domain of ACE requires careful consideration of the small structural differences in the respective domain active sites, such as the Arg381 to Glu403 and Tyr369 to Phe391 substitutions. This objective warrants a structure-based approach.

Many high-resolution crystal structures detailing the interactions between this target and a myriad of different inhibitors have now been reported.[9, 13] A combination of crystal structure data, a large collection of ACEi from decades of ACEi development, modern computer hardware and state-of-the-art docking software have provided an ideal opportunity to develop novel N-domain selective ACEi.

In the search for drug-like N-domain selective ACEi, a large library of known ACEi were screened for a versatile P_2 group with the potential to probe the structure-activity-relationship (SAR) around the residue substitutions found in the S_2 subsite. Here we report an integrated approach to the identification of a novel drug-like N-domain selective ACEi. The approach entailed database mining, docking simulations, hit identification, synthesis, biological evaluation and high-resolution crystal structure studies.

RESULTS AND DISCUSSION

In an effort to identify new starting point leads for a novel N-domain selective ACEi, we screened the comprehensive Gunupati Venkata Krishna (GVK) Biosciences database, containing inhibitors against most biological targets and their associated SAR data.[14] The GVK database contained 24 958 compounds with data from biological studies performed on human sACE. Within this set, 1 832 unique compounds were found to contain a carboxylic acid ZBG with IC_{50} inhibition data.

Manual filtering was used to search for polar P_1/P_2 moieties. Within this carboxylic acid ZBG subset, 13 compounds (**1-13**; Table 1) fit the filtering criteria of a polar group in the P_1/P_2 position. This set of compounds contained four chemical series (Series 1-4) defined by a shared chemical scaffold. Compounds not belonging to one of these scaffolds were left ungrouped. Once these promising compounds had been identified, their binding pose was predicted using constrained Glide docking.[15]

The ability of Glide to evaluate metal-ligand interactions is inadequate for the accurate prediction of an ACEi binding pose. This has been previously reported[16] and is due to the prohibitively expensive quantum mechanical (QM) calculation required to accurately model the force field around transition metals. Consequently, Glide approximates the chelation interaction between a metal and a ligand using a template of pre-calculated interactions between each metal and its chelating species. Such an approximation does not account for the unique electrostatic environment of each metal inside each protein. The use of a pre-calculated generalised template for each metal-chelator interaction is problematic with ligands containing multiple metal-binding moieties of the same species. Most ACEi compounds fall into this category with two carboxylic acids. The use of chelation templates often fail to assign the correct metal binding group when presented with a choice of two carboxylic acids. Hence Glide could not efficiently recreate the native pose of a metal-chelating ACEi in the binding site without the use of docking constraints.

In this case, five docking constraints were created in a docking grid centred around the Lisinopril ligand in the C-domain structure 1O86[9] and RXP407 in the N-domain structure 3NXQ.[13] These grids were assigned three positional constraints requiring the docked ligand to place either a carbon or nitrogen atom in one of three 1 Å radius spheres indicated in Figure 2A. Interaction constraints were then created for a chelation interaction between the ZBG and the zinc metal (Figure 2B). H-bond constraints were set between the terminal carboxylic acid and the H-bond donor Lys and Gln residues in the S_2' subsite (Figure 2C).

Three different constraint conditions were created from different permutations of these five docking constraints, C1, C2 and C3 (Table 2). Different constraint conditions were employed due to the different structural requirements of each chemical series. In each case, one of the three constraint conditions enabled Glide to assign the correct ZBG and find a plausible docking pose (Table 2).

All compounds were successfully docked with a plausible pose into at least one of the two active sites using one of the three constraint conditions (Table 3). These details were then used to assess the potential of a compound to selectively inhibit the N-domain of ACE. This assessment of domain selectivity initially considered the distance from the ligand to a N atom from Arg381 and an O from Tyr369 in the N-domain, and the nearest aromatic C atom from Phe391 and an O atom of Glu403 in the C-domain.

Figure 3 shows example representations of three compounds docked into the N- and C-domains. Compound **1** of Series 1 displayed the ideal interactions with Tyr369 and Arg381 of the N-domain in the S_2 subsite. The other two compounds in this series demonstrated the versatility of this series with the modification to its P_2 group occurring in the last stage of its synthesis. Variability in the P_2 position is ideal for probing the SAR of the S_2 subsite and possible causes of N-domain selective ACE inhibition. Series 2 and 3 provided limited opportunities for probing S_2 interactions apart from a basic Lys group. Large bicyclic groups in the P_2' position have also been shown to cause C-domain selectivity.[17] Series 4 consisted of a pair of natural products with a promising P_1 acidic group. Despite the strong sACE inhibition, these two compounds appeared to be too short to make strong contacts with key S_2 residues that have been shown to influence N-domain selectivity, although a weak, longer range H-bond with Tyr369 is possible. The combination of the shorter length and limited synthetic versatility of a natural product left little opportunity to explore the P_1/P_2 SAR of ACE using this series. An evaluation of the binding poses of the remaining three ungrouped compounds showed little opportunity for P_1/P_2 SAR studies.

From the docking results, the diprolyl Series 1 with its variable P_2 moiety, published by Greenlee et al.[18] appeared to be an ideal scaffold to make contact with key residues in the S_2 subsite. In addition, the P_2 acid found in **1** made it an ideal candidate for N-domain selectivity as it docked with a salt-bridge interaction with Arg381 and formed an H-bond with Tyr369, while its prime groups made the conventional interactions of an ACEi scaffold. Upon closer inspection, **1** and **3** were found to differ from the respective referenced compounds of **14** and **15** published by Greenlee et al. (Table 4). Despite the error present in the GVK database, the diprolyl series presents a promising scaffold with which to probe the P_2 SAR as a large range of amino acids and amino acid analogues can be added in the final stage of the synthesis via the formation of an amide bond.

Greenlee et al. synthesised seven compounds in the diprolyl series. Of the seven compounds, three (**2**, **14** and **15**) contained a terminal P_2 moiety which was free to interact with the Arg381/Glu403 and Try369/Phe391 residues in the N- and C-domains, respectively. Since the P_2 moiety can be replaced with any amino acid or monomer containing a carboxylic acid during the synthesis, a library of P_2 diprolyl variants was modelled.

The diprolyl series was expanded with a diverse range of amino acids and monomer carboxylic acid containing groups (Table 5). These groups were chosen for their ability to probe the P_2 SAR with a carboxylic acid in a variety of positions and orientations. These compounds were docked using the fixed core method. The backbone was fixed in the most plausible pose from the first round of docking with flexibility only permitted in the P_2 position. The resulting poses were then minimised. The final poses can be easily visualised in the 2D ligplot diagrams provided in SI1.

When reviewing the docked poses, the position of the P_2 group was primarily considered due to the rest of the diprolyl scaffold having conserved interactions. The distance between this

group and the Tyr369/Arg381 and Phe391/Glu403 residues in the respective N- and C-domains were measured. Favourable H-bond and salt-bridge interactions with Tyr369/Arg381 were noted, and any additional interactions, both favourable and unfavourable towards the P_2 group were considered.

Figure 4 shows example representations of 3 compounds docked into the N- and C-domains. The P_2 group of compound **16** shows a network of interactions consisting of a salt bridge with Arg381 and H-bonds with Tyr369 and Arg500 (weaker at 3.4Å) in the N-domain. In contrast, when docked into the C-domain, the charged P_2 group rotates away from the hydrophobic Phe391 and only forms a single interaction, albeit a strong H-bond with Arg522, the equivalent residue to Arg500 in the N-domain. This movement of the P_2 group of compound **16** between the N- and C-domains resulted in orientation towards the S_3 subsite and is typical of compounds that were either too long for the S_2 subsite, or were not able to form favourable interactions with Tyr369/Phe391 and/or Arg381/Glu403.

The P_2 group of compound **17** is one alkyl carbon longer than that of **16.** The most obvious effect of this difference is the change, for both the N- and C-domains, from a trans proline in the P_1 position of **16** to a cis proline in **17**. In the trans orientation, the P_2 backbone carbonyl interacts with the backbone nitrogen of Ala334 and Ala356 for the N- and C-domains respectively. In contrast, with a cis proline the P_2 backbone carbonyl interacts with Thr496 in the N-domain, which is not conserved in the C-domain, the interaction is instead with Arg522. This cis proline is also observed for **22** with both the N- and C-domains, **19** with the N-domain, and **2**, **3** and **15** with the C-domain. For the N-domain, the change to a cis proline with **17** only has small effect on the position of the P_2 carboxyl group, but the orientation indicates potentially stronger interactions with Tyr369 and Arg381. In addition, there is a large change in the position of the P_2 backbone nitrogen from interacting with the backbone carbonyl of Ala334 for **16**, to H-bonds with Glu389 and Tyr501 for **17**. A comparison of **16** and **17** with the C-domain shows a shift of the P_2 carboxyl group closer to Phe391 for the docked **17**, and this causes the loss of interaction with Arg522 seen for **16**. Similar to the N-domain, there is a large change in position of the P_2 backbone nitrogen between **16** and **17** in the C-domain, which interacts with Ala356 for **16**, but Glu411 for **17**.

Compound **21** is an example where the P_2 group is not an amino acid, and possibly a consequence of this is that the P_1 proline backbone is twisted in both the N- and C-domains, therefore neither cis nor trans. In the N-domain, the P_2 carboxyl group forms the desired interactions with Tyr369 and Arg381, whereas in the C-domain this group twists away from the hydrophobic Phe391, but does form an H-bond with Arg522. There is also a difference with the P_2 backbone carbonyl which interacts with Thr496 in the N-domain (not conserved in the C-domain), but Tyr523 in the C-domain.

In addition to the qualitative assessment of the binding pose, a quantitative molecular mechanics – generalised Born surface area (MM-GBSA) binding energy calculation was performed to provide a metric with which to assess the binding prediction. The combination of the qualitative binding assessment and the quantitative MM-GBSA binding energy calculation for each ligand in each domain helped to form a binding prediction.

Upon the formulation of the binding predictions, all compounds were deemed to be specific for either the N-domain (N) or C-domain (C), or non-selective (NS) (Table 5). Considering these selectivity predictions and availability of starting materials, a representative set of seven compounds

(indicated with an asterisk in Table 5) was selected for synthesis. These compounds were chosen for proof-of concept studies to test the hypothesis that an acid group in the P_2 position is crucial for N-domain selective ACE inhibition.

Chemistry.

The diprolyl series was synthesised in 8-steps as depicted in scheme 1. For reaction i *N*-Boc pyrrolidine aldehyde was mixed with Ala-Pro in MeOH/$NH_4^+Cl^-$ solution at 25 °C for 36 hours to produce **27** with a yield of 36%. This compound was then oxidised in an acidic HCl/MeOH solution at 25 °C for two days. After a quick work-up the product was dissolved in a solution of $(Boc)_2O$ and Et_3N/MeOH to reprotect any Boc-amine groups hydrolysed during a side reaction to produce **28** (72%) which was then loaded onto an Amberlyst 15 resin and stirred in MeOH at 60 °C for 10 days. Compound **29** was then eluted from the resin using HCl in dioxane (40%). The RCO_2H group was then attached to **29** using T3P in a basic Et_3N solution at 0-25 °C for 12 hours to form **30_x** with **x** denoting the final compound to be synthesised after its deprotection (**30_2** was deprotected into compound **2**). The product was then de-esterified using LiOH in THF/H_2O and Boc deprotected using HCl/Dioxane if the compound still contained a Boc protected amine.

Biological Evaluation.

Docking and MM-GBSA simulations predicted compounds **2** and **14** to be strong inhibitors of both ACE domains, **15** to potentially have some C-domain specificity, while **3**, **16**, **17** and **18** were predicted to show a degree of selectivity towards the N-domain. Only **3**, **16** and **17** met both the qualitative and quantitative criteria for selectivity, albeit the difference between the MM-GBSA binding energy for the N- and C-domains for **3** was small.

The K_i values of all seven synthesised diprolyl compounds were evaluated against both domains via a fluorimetric competitive inhibition assay (Table 6). All compounds inhibited both domains of ACE in the nanomolar range. Compounds **3**, **14** and **18** showed very low nanomolar inhibition of both domains with a small degree of C-domain selectivity. Of the four compounds predicted to selectively inhibit the N-domain of ACE, only **16** was specific for the N-domain with a similar affinity as **3**, **14**, and **18**, but an 84-fold lower affinity for the C-domain. In contrast, compound **17** had the same high affinity for both domains showing no specificity. Compounds **2** and **15** were lower affinity inhibitors for both domains, with **2** showing a small level of C-domain specificity.

The design and docking procedures have produced high affinity inhibitors of ACE. The data also shows that introducing the P_2 carboxylic acid group can give the desired N-domain specificity as seen with compound **16**. It is not immediately obvious from the docking results why compound **17** has a higher affinity for the C-domain than **16**, and therefore why **17** loses N-domain specificity. Compound 17 docked with the cis proline conformation. Although this causes the loss of the H-bond with Arg522, and there are no additional interactions, the cis proline conformation may allow the conserved interactions to be stronger. Further SAR and structural studies are needed to fully understand these subtle differences.

X-ray Crystallography.

Three structures (A and B) of the N-domain of ACE in complex with compound **16** are presented in this paper (Table 7), both of which crystallised in the usual P1 space group previously seen for the N-domain of ACE. These two structures have a larger unit cell (Table 7) than what is typical of the ACE N-domain with 4 chains in the asymmetric unit, similar to the structure of the N-domain in complex with amyloid-beta 4-10 (PDB code 5am8). However, the overall structure of all chains in the unit cells of both structures are similar to each other and previously reported structures. Residues 10-23 and 74-90 form part of the proposed N-terminal hinge region and often show flexibility in crystal structures with increased temperature factors (B-factors) and poor electron density. This is seen to some degree in every chain of both reported structures, but to a greater extent in chain B of each structure. In addition, chain B of structure A shows a particularly high degree of flexibility with some evidence of a second conformation apparent in the mFo-DFc electron density map of this region. The electron density for this second conformation is not sufficient to model the structure, but to account for this the first conformation of this region has been assigned a 0.8 occupancy. Another difference between the chains in the asymmetric unit is that an extra 7 (structure A) and 8 (structure B) residues can be resolved in Chain D at the C-terminus due to the proximity and interaction with a symmetry-related molecule. In both structures, there is a region of positive density in the mFo-DFc difference map adjacent to residue Asp177 of chain D. This cannot be properly accounted for by waters or any other constituent of the purification or crystallisation conditions, and has therefore been left un-modelled. A further significant difference between the molecules, is seen with Arg381 as described in detail below.

The active site of the N-domain/compound **16** complex structure shows clear electron density for the ligand in every molecule of the asymmetric unit in both structures (Figure 5A). Compound **16** binds to the active site in the expected position and orientation coordinating the zinc metal with its carboxylic acid ZBG, while the remainder of the ligand occupies the S_2', S_1', S_1 and S_2 subsites. The C-terminus mimic of **16** occupies the usual position in the S_2' subsite forming strong hydrogen bonding and salt bridge interactions with Gln259, Lys489 and Tyr498 (Figure 5B). These residues are conserved in the C-domain. The P_1' backbone carbonyl forms hydrogen bonds with the His331 and His491 sidechains, and the P_1' backbone nitrogen with the backbone carbonyl of Ala332. These residues are also conserved in the C-domain and the interactions are typical in most structures of the N-and C-domains in complex with different ligands. The ZBG of **16** also interacts with the C-domain conserved residues Glu362 and Tyr501. The S_1 subsite contains the Thr496 to Val518 substitution between the N- and C-domains, making the S_1 subsite more hydrophobic in the C-domain. This would most likely favour the pseudo-proline P_1 group of **16**, and therefore cannot account for the difference in affinity towards **16**.

The interactions between **16** and the S_2', S_1' and S_1 subsites involve largely conserved residues. The observed specificity for the N-domain must therefore arise from the interactions between the P_2 group and residues in the S_2 subsite. Since the binding site cavity begins to widen near the S_2 subsite, there are numerous residues that could interact with the ligand depending on the orientation and length of the P_2 moiety. As already mentioned, Tyr369 and Arg381 in the N-domain are the only two residues not conserved in the C-domain that could realistically provide different interactions with **16**. These residues are replaced by Phe391 and Glu403 respectively in analogous positions in the C-domain. As previously mentioned, interactions with Tyr369 and Arg381 were targeted due to the interactions observed with RXP407 in the 3NXQ N-domain complex structure. In all the molecules of the asymmetric unit for both structures, Arg381 is poorly defined

beyond the Cα (Figure 6A-C). The weak density suggests a high degree of flexibility, with two orientations partly explaining some of this flexibility. However, neither conformation interacts directly with compound **16**. There is a varying amount of electron density visible for these conformations in the different chains, but each conformation is visible to some extent in at least one chain of both structures. This collective observation across the 8 chains was used to model these alternative conformations into weak density with a high degree of certainty (a true experimentally observed '*dynamic picture'* emerged from the crystal structures with multiple copies in the asymmetric unit of the crystal lattice), and is the reason for including 2 versions of the structure. One of these conformations allows for a water mediated interaction between Arg381 and the P_2 carboxylic acid, with the water bridging both oxygens of the carboxylic acid. However, due to the high degree of flexibility of Arg381 observed in these structures, it is unlikely that interactions with this residue significantly contributes to the N-domain specificity of **16**. The second conformation orientates away from the bound ligand, and there is variation in this conformation across the different chains in the asymmetric units further confirming the flexibility of this residue.

In contrast to Arg381, the crystal structure shows a strong H-bond interaction between the P_2 carboxylic acid of **16** and the hydroxyl group of Tyr369 (Figure 6D). The equivalent Phe391 residue in the C-domain, lacks the hydroxyl hydrogen bond donor group while creating a hydrophobic environment that is not suitable for charged moieties. Therefore, it is likely that the interaction between **16** and Tyr369 is a major contributing factor for the N-domain specificity of this compound.

The P_2 carboxylic acid also has water mediated interactions with Arg500 (Figure 6D) contributing to the high affinity of **16** for the N-domain. The docking results predicted this as a long range (3.4 Å) direct interaction, and an H-bond between **16** with the equivalent residue Arg522 of the C-domain. Therefore, the interaction seen with Arg500 of the N-domain may not contribute to the specificity of **16**.

Comparison of Crystallography and Docking Results.

The crystal structure of compound **16** in the N-domain overlays very closely with the predicted docked pose (Figure 7A). An RMSD of 1.182 Å was measured between the predicted and crystallised poses. The diprolyl scaffold shows a strong agreement with just a small difference in the orientation, but not the position of the P_2 aspartate. This close overlay strongly validates the constrained docking methodology. The only notable difference between the predicted and crystallised structures was the lack of a salt-bridge between the P_2 carboxylic acid and Arg381. This is a consequence of docking the ligand against the crystal structure of 3NXQ where Arg381 is orientated inwards to make a salt-bridge contact with the P_2 carboxylic acid of the larger RXP407 ligand. The choice of the 3NXQ structure is justified for this purpose to maximise the identification of compounds that could potentially interact with Arg381

Overlaying the docked C-domain pose of **16** with the N-domain crystal structure (Figure 7B) illustrates the change in orientation of the P_2 group that is possibly required to avoid an unfavourable interaction with Phe391 in the C-domain. This may also be driven by the favourable direct interaction with Arg522 that resulted from this change. In the N-domain crystal structure, the interaction with the equivalent Arg500 is water mediated.

The docking results for **16** are therefore consistent with the results from the crystal structure and the biological assay with **16** having a greater number of interactions with the N-domain than with the C-domain, particularly the strong H-bond with Tyr369, thereby explaining the specificity for the N-domain.

Comparison of N-domain Specific Inhibitor Structures.

The work presented here investigated whether N-domain specificity could be achieved by targeting interactions with Tyr369 and Arg381 in the N-domain S_2 subsite. These residues were identified for their important interactions contributing to the N-domain specificity of RXP407.[13] A comparison of crystal structures of the N-domain in complex with compound **16** and RXP407 (depicted as an overlay in Figure 8A), show the compounds occupying very similar positions with the majority of interactions conserved. In the S_2 subsite, **16** retains the Tyr369 interaction, but Arg381 at best only provides a transient water mediated interaction. This is consistent with the 2500-fold N-domain selectivity of RXP407 compared with compound **16** showing an 84-fold N-domain selectivity. Compared to RXP407, the P_2 carboxylic acid of **16** is shifted approximately 0.5 Å away from the Arg381 location due to differences in ligand structure. This small shift appears to be sufficient to prevent a strong interaction with Arg381 indicating the sensitivity of the binding to slight changes in ligand structure.

33RE is another N-domain specific inhibitor, with identical groups to RXP407 in the P_2', P_1' and P_1 positions, but with the P_2 group including the substitution of the carboxylic acid for a tetrazole group. 33RE has a 1000-fold specificity towards the N-domain. An overlay of the 33RE structure[19] with that of **16** shows the P_2 groups occupy a similar position (Figure 8B). The tetrazole of 33RE interacts with Tyr369, similar to both RXP407 and **16**, but without a direct interaction with Arg381. It is interesting to note that due to this lack of interaction, the Arg381 in the 33RE structure overlays with the second conformation of Arg381 in the **16**-complex structure. Comparison of the P_2 group interaction does not explain why 33RE shows a greater specificity for the N-domain than observed with **16**.

Previous analysis of N-domain specificity shown by RXP407 and 33RE[20] indicated that a P_2' carboxylic acid in place of an amide group greatly reduces, or even abolishes the N-domain specificity. Both RXP407 and 33RE contain P_2' amide groups, whereas **16** has a carboxylic acid. It is possible that changing this P_2' carboxylic acid of **16** to an amide may greatly increase the specificity for the N-domain. In addition, it is conceivable that compounds **3, 17**, and **18**, predicted to be N-domain selective, but shown to be non-specific or even having some C-domain specificity, are having their P_2 N-domain selectivity masked by the P_2' carboxylic acid. Synthesis of P_2' amide analogues of these compounds would therefore be of considerable interest.

Conclusion.

We have investigated a series of diprolyl compounds for specificity towards the N-domain catalytic site of ACE by screening the GVK database for potentially N-domain selective P_2 chemical groups. The series was expanded with a diverse range of amino acids and monomer carboxylic acid containing groups to explore the S_2 subsite. These compounds were assessed using molecular docking and MM-GBSA to estimate their relative binding affinity. Nine compounds with different P_2 groups were predicted to bind selectively to the N-domain. A representative set of seven

compounds, four of which were predicted to be N-domain selective, 1 C-domain specific and 2 non-specific, were synthesised using an 8-step synthetic route. While six of the compounds showed highly potent N- and C-domain inhibition, only the P_2 Asp derivative (**16**) was N-domain-selective and bound with an 84-fold higher affinity than the C-domain.

The high-resolution crystal structures of compound **16** in complex with the N-domain showed that the ligand position, orientation and interactions largely agreed with the docking results. The only significant difference was the lack of a salt bridge with Arg381. This residue was shown to be highly flexible forming only a transient water mediated interaction with the with the P_2 carboxylic acid. Instead, hydrogen bonding of **16** with Tyr369 is likely to be the interaction responsible for N-domain selectivity. The close correlation between the predicted and crystallised pose of **16** reaffirms accuracy of the pose prediction while the movement of Arg381 highlights the shortcomings of a fixed receptor prediction.

The discovery of the diprolyl series has provided a more drug-like scaffold with which to probe the S_2 subsite of ACE for N-domain selectivity. While the discovery of **16**, a novel N-domain selective ACEi is encouraging, it highlights the need for a more extensive SAR study probing the S_2 subsite to better define the chemical space responsible for N-domain selective ACE inhibition. A previous study showed a complete loss in N-domain specificity caused by a change from a P_2' amide to carboxylic acid. That SAR masked the contribution of the P_2 group.[12] The present study has identified the first drug-like N-domain selective ACEi. Compound **16** serves as an excellent starting point for further SAR studies which could realistically develop into a viable N-domain selective ACEi and achieve the original goal of treating tissue fibrosis without affecting blood pressure.

METHODS

Computational.

All computational predictions were carried out using desktop workstations running the CentOS 6.5 OS using the Glide utility included in the Schrodinger 2014-1 suite release. PDB entries 3NXQ and 1O86 of the respective N- and C-domains were prepared using the Maestro prepwizard to complete the structure with the addition of bond orders and missing side-chains. The non-peptide ligands were edited manually to reflect their correct structure. Non-bound waters were removed and in the case of 3NXQ, the less complete Chain-B was removed. The automated optimize and Impref protocols were then run to refine the structures. Glide docking grids of default length were created centred on the native Lisinopril and RXP407 ligands, each with one metal chelation constraint, three positional constraints and two S_2' H-bond constraints as indicated in Figure 3. Docking simulations were performed iteratively using the Glide SP setting with constraint conditions CS1-CS3 enforced until a plausible docking pose was found.

Binding energy was calculated using the Prime MM-GBSA minimisation and binding energy calculation package provided with the Schrodinger suite. MM-GBSA calculations were performed using the VSGB solvent model. The minimisation was performed with flexibility tolerated for all protein atoms within a 15 Å radius of the ligand.

Synthesis

Reagents and Solvents.

All chemical reagents and anhydrous solvents used during the synthesis of the Enalaprilat analogue were purchased from Sigma Aldrich. HPLC grade solvents were purchased from Sigma-Aldrich, Merck and Microsep

Chromatography.

Thin layer chromatography (TLC) reaction monitoring was performed using Merck F_{254} aluminium-backed silica gel 60 plates. Spots were visualised with either ultra violet (UV) light (254/366 nm), anisaldehyde or ninhydrin stains. Reaction products were all purified via column chromatography using Merck kieselgel 60:70-230 mesh via gravitational column chromatography and flash chromatography.

HPLC.

Peak purity of each compound was determined using preparatory HPLC with a thermo separation system comprising of a Spectra Series P200 pump, an AS100 automated sampler and a UV 100 variable wave detector. The UV detector was set to monitor peak absorption at 214 nm. A Waters® X-bridge C18 5.0 µm (4.6 x 150 mm) (Phenomenex, Torrance, CA) column stationary phase used was fitted to a Supelguard ® Ascentis ™ C18 guardcartridge (2cm x 40 mm, 3 µm) (Supelco Analytical, Bellefonto, PA).

Diastereomer separation was performed using a thermo separation system comprising of an Analytical Technologies® P2230 HPLC pump with an automated sampler and Analytical Technologies® UV 2230 variable wavelength detector set to detect absorbance at 214 nm. The stationary phase column used was a Gemini® NX – C18 3.5 µm (4.6 x 50 mm) (Phenomenex, Torrance, CA).

LC-MS.

Liquid chromatography mass spectrometer (LC-MS) analysis on each compound was performed using an Agilent® 1260 Infinity Binary Pump, Agilent® 1260 Infinity Diode Array Detector (DAD), Agilent® 1290 Infinity Column Compartment, Agilent® 1260 Infinity Standard autosampler and an Agilent® 6120 Quadruple (Single) mass spectrometer equipped with an APCI and ESI multimode ionisation source. Compound purity was determined using an Agilent® LC-MS with a Kinetex Core C18 2.6 µm column (50 x 3 mm).

NMR.

All NMR spectra were recorded a Brucker Ultrashield-Plus Spectrometer (^{1}H-400 MHz; ^{13}C-100 MHz), with compounds dissolved in either deuterated methanol (MeOD-d_4) or deuterium oxide (D_2O) solvents.

Inhibitions Assays.

The Ndom389 and tACE Δ36NJ constructs of the respective N- and C-domain samples were diluted to 10 nM and 5 nM respectively from frozen stock solutions. 50 mM stock solutions of each compound were prepared in deionised water. Aliquots of these stock solutions were then diluted to 10 mM with deionised H_2O followed by dilution into a phosphate buffer (100 mM $KHPO_4$, pH 8.3, 300 mM NaCl, 10 µM $ZnSO_4$, 1mg/ml albumin).

10-fold dilution series running from 100 µM to 1 nM of each compound were prepared in buffer. 40 µL of each solution in the series was mixed with 40 µL of the enzyme in buffer. A 10 µM solution of Lisinopril in buffer was used for positive control while buffer was used for the negative control. The mixtures were incubated at room temperature for 30 minutes after which triplicate 20 µL aliquots of each mixture were loaded into 3 individual wells of a 96 well plate. 30 µL of 1 mM Z-FHL substrate (Bachem Ltd., Bubenhof, Switzerland) was then added to each well. The plate was then incubated at 40 ˚C for 15 minutes in a shaker. The reactions were then stopped and the product was derivatised with the addition of 190 µL of a base and *O*-pthaldehyde solution (0.28 M NaOH and 7 mM *O*-pthaldehyde) to each well followed by further incubation on a shaker at room temperature for 10 minutes. The wells were then quenched with 25 µL of 3 M HCl and the plate was read on a fluorimeter (Varian Inc., Mulgrave, Victoria, Australia) with an excitation wavelength at 360 nm and an emission wavelength at 485 nm. After approximating the IC_{50} values from the 10-fold dilution series of each compound, 9 solutions of a two-fold dilution series spanning the approximated IC_{50} were prepared. Activity curves were plotted to determine the IC_{50} of each compound. For better inhibition comparisons between the two domains of ACE, the K_i values were calculated using the equation given below:

$$K_i = \frac{IC_{50}}{1 - \frac{[S]}{K_m}}$$

Characterisation

General Ester Deprotection Method.

5 equivalents of a 4.0 M aqueous solution of LiOH was added to a 0.1 M solution of the methyl ester protected form of each diprolyl compound in a THF/MeOH/H_2O (3:2:1) solution. The mixture was stirred for two hours at room temperature. The mixture was then diluted with 5 mL of water then neutralised by adding 1 N HCl dropwise. The compound was either dried and purified via Prep-HPLC or the crude product was Boc-deprotected.

General Boc Deprotection Method.

Crude Boc protected compound was dissolved in chilled 4 M HCl/Dioxane under N_2 at 0 °C. The mixture was stirred for 30 min at 0° C and monitored via TLC. Upon completion, the solvent was evaporated and the residue was washed with diethyl ether. The final compound was then purified over prep HPLC.

2

The general ester deprotection was performed on 90 mg of **30_2** after which the crude residue was Boc deprotected via the general Boc deprotection method after which it was purified via HPLC to yield 65 mg (97%) of **30_2** as a sticky yellow solid.

^{1}H NMR (400 MHz, D_2O) δ 4.42 (m, 2H), 4.29 (q, *J* = 6.9 Hz), 3.95 – 3.84 (m, 3H), 3.54 (m, 2H), 3.48 – 3.33 (m, 2H), 2.25 (s, 1H), 2.15 (m, 1H), 2.04 (m, 1H), 1.95 (m, 4H), 1.82 (m, 1H), 1.50 (d, *J* = 6.9 Hz, 3H). ^{13}C (100 MHz, D_2O): δ 177.1, 176.2, 174.7, 170.9, 70.5, 63.1, 56.3, 55.5, 52.3, 51.5, 42.9, 31.8, 26.9, 25.9, 24.1, 23.5 LC-ESI-MS (+ve ion mode): 371 $[M+H]^+$ HPLC purity 93.6%, t_r = 5.66 min

3

The general ester deprotection was performed on 50 mg of **30_3** after which the crude residue was Boc deprotected via the general Boc deprotection method after which it was purified via Prep HPLC to yield 30 mg (78%) of **30_3** as a sticky yellow solid.

^{1}H NMR (400 MHz, D_2O) δ 7.61 – 7.51 (m, 2H), 7.27 (dd, *J* = 8.7, 2.9 Hz, 2H), 4.51 (t, *J* = 8.5 Hz, 1H), 4.33 (q, J = 6.9 Hz, 1H), 4.29 (m, 1H), 3.96 (m, 1H), 3.62 – 3.36 (m, 4H), 2.21 (m, 2H), 2.13 – 2.00 (m, 1H), 1.98 – 1.79 (m, 4H), 1.66 (m, 1H), 1.50 (d, *J* = 6.9 Hz, 3H). ^{13}C (100 MHz, D_2O): δ 176.3, 174.8, 174.1, 168.3, 133.9, 128.1, 127.7 (2C), 126.8 (2C), 67.5, 66.1, 60.7, 58.2, 52.7, 48.9, 28.1, 26.4, 23.7, 22.1, 21.1. LC-ESI-MS (+ve ion mode): 433 $[M+H]^+$ HPLC purity 93.79%, t_r = 4.95 min

14

The general ester deprotection was performed on 45 mg of **30_14** and then purified via prep HPLC to obtain 26 mg (62%) of a sticky white-yellow solid.

^{1}H NMR (400 MHz, Methanol-d_4) δ 7.62 – 7.55 (m, 2H,), 7.55 – 7.44 (m, 3H, H-12), 4.72 (t, *J* = 8.7 Hz, 1H), 4.52 – 4.47 (m, 1H), 4.37, (d, J = 1.9 Hz, 1H), 4.23 (q, J = 6.8 Hz, 1H), 3.70 (m, 1H), 3.60 (m, 2H), 3.50 (d, 1H), 2.44 – 2.25 (m, 3H, H-2), 2.12 – 1.93 (m, 4H), 1.87-1.73 (m, 1H), 1.64 (d, *J* = 7.0 Hz, 3H). ^{13}C (100 MHz, Methanol-d_4): δ 175.9, 175.1, 173.2, 168.2, 132.3, 129.4, 127.9 (2C), 126.5 (2C), 66.3, 64.2, 59.5, 58.3, 50.5, 49.3, 27.2, 25.3, 23.4, 22.7, 22.1 LC-ESI-MS (+ve ion mode): 418 $[M+H]^+$ HPLC purity 98%, t_r = 8.26 min

15

The general ester deprotection was performed on 50 mg of **30_15** after which the crude residue was Boc deprotected via the general Boc deprotection method after which it was purified vi prep-HPLC to yield 38 mg (97%) of **30_15** as a sticky yellow solid.

^{1}H NMR (400 MHz, D_2O) δ 7.32 (m, 3H), 7.25 (m, 2H), 4.51 (t, J = 7.6 Hz, 1H), 4.40-4.30 (m, 3H, H-4), 3.82 (d, *J* = 4.4 Hz, 1H), 3.64 (d, 2H), 3.54 – 3.39 (m, 2H), 3.23 (dd, *J* = 14.7, 5.1 Hz, 1H), 3.06 (dd, *J* = 14.8, 7.7 Hz, 1H), 2.21 (m, 1H), 2.07 (m, 1H), 2.00 – 1.81 (m, 6H), 1.48 – 1.36 (m, 3H). ^{13}C (100 MHz, D_2O): δ 177.1, 175.3, 174.8, 171.2, 136.7, 129.1 (2C), 128.5 (2C), 127.2, 68.3, 67.5, 62.1, 57.3, 53.6, 51.9, 49.2, 40.1, 29.3, 25.7, 24.1, 22.8, 22.1 LC-ESI-MS (+ve ion mode): 461 $[M+H]^+$ HPLC purity 97.4%, t_r = 5.52 min

16

The general ester deprotection was performed on 230 mg of **30_16** after which the crude residue was Boc deprotected via the general Boc deprotection method after which it was purified via prep-HPLC to yield 85 mg (49%) of **30_16** as a sticky yellow-brown solid.

^{1}H NMR (400 MHz, D_2O) δ 4.60-4.50 (m, 1H), 4.48-4.35 (m, 2H), 4.33-4.24 (m, 1H), 3.85 (d, *J* = 2.1 Hz), 3.70-3.48 (m, 3H), 3.47-3.35 (m, 1H), 3.03-2.91 (m, 1H), 2.84 – 2.71 (m, 1H), 2.23-2.14 (m, 2H), 2.13 – 1.75 (m, 6H), 1.59 – 1.35 (d, *J* = 6.1 Hz, 3H). ^{13}C (100 MHz, D_2O): δ 177.1, 176.5, 175.9, 174.2 170.6, 70.9, 68.5, 59.5, 58.2, 53.8, 50.2, 49.1, 41.9, 31.1, 28.3, 27.1, 24.1, 23.5. LC-ESI-MS (+ve ion mode): 429 $[M+H]^+$ HPLC purity 97.55%, t_r = 4.76 min

17

The general ester deprotection was performed on 80 mg of **30_17** after which the crude residue was Boc deprotected via the general Boc deprotection method after which it was purified to yield 6 mg (10%) of **30_17** as a sticky yellow-brown solid.

^{1}H NMR (400 MHz, D_2O) δ 4.40 – 4.23 (m, 4H), 3.76 (d, *J* = 2.0 Hz, 1H), 3.65 (m, 1H), 3.53 (m, 2H), 3.39 (m, 1H), 2.54 (t, *J* = 7.5 Hz, 2H), 2.22 (m, 2H), 2.13 (m 2H), 1.95 (m, 5H), 1.77 (q, 1H), 1.49 (d, *J* = 6.9 Hz, 3H). ^{13}C (100 MHz, D_2O): δ 178.2, 177.4, 175.5, 174.9 171.4, 72.3, 68.9, 61.5, 58.1, 51.9, 51.2, 48.3, 41.9, 31.1, 29.7, 28.1, 26.4, 24.2, 22.9. LC-ESI-MS (+ve ion mode): 443 $[M+H]^+$ HPLC purity 98.95%, t_r = 4.76 min

18

The general ester deprotection was performed on 30 mg of **30_18** and then purified via prep HPLC to obtain 23 mg (62%) of a sticky white-yellow solid.

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.07 (d, *J* = 7.8 Hz, 2H), 7.65 (dd, *J* = 8.2, 2.9 Hz, 2H), 4.71 – 4.61 (t, J = 8.2 Hz), 4.40 (q, *J* = 6.9 Hz, 1H), 4.12 (m, 1H), 3.96 (m, 1H), 3.62 (m, 2H), 3.43 (m, 2H), 2.32 (m, 3H), 1.98 (m, 3H), 1.86 (m, 1H), 1.74 (m, 1H), 1.58 (d, *J* = 6.4 Hz, 3H). ^{13}C (100 MHz, Methanol-d_4): δ 175.9, 175.1, 174.5, 171.3, 167.2, 134.3, 128.8, 128.1 (2C), 126.3 (2C), 67.1, 66.2, 59.3, 57.9, 51.5, 49.1, 27.3, 25.8, 23.1, 22.5, 21.3. LC-ESI-MS (+ve ion mode): 462 $[M+H]^+$ HPLC purity 97%, t_r = 7.31 min

Protein Purification and X-ray Crystallography.

The N-domain of ACE was expressed and purified to homogeneity from CHO (Chinese-hamster ovary) cells.[13] Compound **16** in complex with Ndom389 (minimally glycosylated ACE N-domain) was prepared by incubating a 4:1 v/v ratio of 5 mg ml^{-1} Ndom389 (in 50 mM Hepes, pH 7.5, 0.1 mM PMSF) and 20 mM **16** at room temperature for 1 hour. The hanging drop co-crystallization used 1 µL of the complex mixed with an equal volume of reservoir solution (30% PEG 550 MME/PEG 20000, 0.1 M Tris/Bicine pH 8.5, and 60 mM divalent cations. Molecular Dimensions Morpheus A9).

X-ray diffraction data were collected from two crystals on station IO4-1 at the Diamond Light Source (Didcot, UK). Crystals were kept at constant temperature (100 K) under the liquid nitrogen jet during data collection. Images were collected using a PILATUS-6M-F detector (Dectris, Switzerland). Raw data images were indexed and integrated with DIALS,[21] and then scaled using AIMLESS[22] from the CCP4 suite.[23] Initial phases were obtained by molecular replacement with PHASER[24] using PDB code 3NXQ[12] as the search model. Further refinement was initially carried out using REFMAC5[25] and then Phenix,[26] with COOT[27] used for rounds of manual model building. Ligand and water molecules were added based on electron density in the Fo-Fc difference map. MolProbity.[28] was used to help validate the structures. Crystallographic data statistics are summarized in Table 7. All figures showing the crystal structures were generated using CCP4mg.[29]

ACKNOWLEDGEMENTS

The University of Cape Town, South African Medical Research Council (SAMRC), National Research Foundation (NRF) Competitive Programme for Rated Researchers, and South African Research Chairs Initiative of the Department of Science and Technology, administered through the NRF are gratefully acknowledged for support. KRA wishes to thank the Medical Research Council (UK) for support (Grant Number MR/M026647/1); the scientists at station IO4-1, Diamond Light Source, Didcot, Oxon (UK), for their support during X-ray diffraction data collection. KRA and EDS also thank the University of Cape Town (South Africa) and University of Bath (UK) respectively for Visiting Professorships.

REFERENCES

1. Cushman, D. W.; Cheung, H.; Sabo, E.; Ondetti, M. Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. *Biochemistry* **1977,** 16, 5484-5491.
2. Patchett, A.; Harris, E.; Tristram, E.; Wyvratt, M.; Wu, M.; Taub, D.; Peterson, E.; Ikeler, T.; Ten Broeke, J.; Payne, L. A new class of angiotensin-converting enzyme inhibitors. *Nature* **1980,** 288, 280-283.
3. Bakhle, Y. Conversion of angiotensin I to angiotensin II by cell-free extracts of dog lung. *Nature* **1968,** 220, 919-921.
4. Byers, L. D.; Wolfenden, R. Binding of the by-product analog benzylsuccinic acid by carboxypeptidase A. *Biochemistry* **1973,** 12, 2070-2078.
5. Fuchs, S.; Xiao, H. D.; Hubert, C.; Michaud, A.; Campbell, D. J.; Adams, J. W.; Capecchi, M. R.; Corvol, P.; Bernstein, K. E. Angiotensin-converting enzyme C-terminal catalytic domain is the main site of angiotensin I cleavage in vivo. *Hypertension* **2008,** 51, 267-274.
6. Lenfant, M.; Wdzieczak-Bakala, J.; Guittet, E.; Prome, J.-C.; Sotty, D.; Frindel, E. Inhibitor of hematopoietic pluripotent stem cell proliferation: purification and determination of its structure. *Proceedings of the National Academy of Sciences* **1989,** 86, 779-782.
7. Robinson, S.; Lenfant, M.; Wdzieczak-Bakala, J.; Melville, J.; Riches, A. The mechanism of action of the tetrapeptide acetyl-N-Ser-Asp-Lys-Pro (AcSDKP) in the control of haematopoietic stem cell proliferation. *Cell proliferation* **1992,** 25, 623-632.
8. Rousseau, A.; Michaud, A.; Chauvet, M.-T.; Lenfant, M.; Corvol, P. The hemoregulatory peptide N-acetyl-Ser-Asp-Lys-Pro is a natural and specific substrate of the N-terminal active site of human angiotensin-converting enzyme. *Journal of Biological Chemistry* **1995,** 270, 3656-3661.
9. Natesh, R.; Schwager, S. L.; Sturrock, E. D.; Acharya, K. R. Crystal structure of the human angiotensin-converting enzyme–lisinopril complex. *Nature* **2003,** 421, 551-554.
10. Kröger, W. L.; Douglas, R. G.; O'Neill, H. G.; Dive, V.; Sturrock, E. D. Investigating the domain specificity of phosphinic inhibitors RXPA380 and RXP407 in angiotensin-converting enzyme. *Biochemistry* **2009,** 48, 8405-8412.
11. Soubrier, F.; Alhenc-Gelas, F.; Hubert, C.; Allegrini, J.; John, M.; Tregear, G.; Corvol, P. Two putative active centers in human angiotensin I-converting enzyme revealed by molecular cloning. *Proceedings of the National Academy of Sciences* **1988,** 85, 9386-9390.
12. Dive, V.; Cotton, J.; Yiotakis, A.; Michaud, A.; Vassiliou, S.; Jiracek, J.; Vazeux, G.; Chauvet, M.-T.; Cuniasse, P.; Corvol, P. RXP 407, a phosphinic peptide, is a potent inhibitor of angiotensin I converting enzyme able to differentiate between its two active sites. *Proceedings of the National Academy of Sciences* **1999,** 96, 4330-4335.
13. Anthony, C. S.; Corradi, H. R.; Schwager, S. L.; Redelinghuys, P.; Georgiadis, D.; Dive, V.; Acharya, K. R.; Sturrock, E. D. The N Domain of Human Angiotensin-I-converting Enzyme: The Role of N Domain Glycosylation and the Crystal Structure in Complex with an N Domain-Specific Phosphinic Inhibitor, RXP407. *Journal of Biological Chemistry* **2010,** 285, 35685-35693.
14. Biosciences, G. Drug Database and Clinical Candidate Database. https://www.gostardb.com/index.jsp
15. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry* **2004,** 47, 1739-1749.
16. Hu, X.; Shelver, W. H. Docking studies of matrix metalloproteinase inhibitors: zinc parameter optimization to improve the binding free energy prediction. *Journal of Molecular Graphics and Modelling* **2003,** 22, 115-126.
17. Georgiadis, D.; Cuniasse, P.; Cotton, J.; Yiotakis, A.; Dive, V. Structural determinants of RXPA380, a potent and highly selective inhibitor of the angiotensin-converting enzyme C-domain. *Biochemistry* **2004,** 43, 8048-8054.

18. Greenlee, W. J.; Allibone, P. L.; Perlow, D. S.; Patchett, A. A.; Ulm, E. H.; Vassil, T. C. Angiotensin-converting enzyme inhibitors: synthesis and biological activity of acyltripeptide analogs of enalapril. *Journal of medicinal chemistry* **1985,** 28, 434-442.
19. Douglas, R. G.; Sharma, R. K.; Masuyer, G.; Lubbe, L.; Zamora, I.; Acharya, K. R.; Chibale, K.; Sturrock, E. D. Fragment-based design for the development of N-domain-selective angiotensin-1-converting enzyme inhibitors. *Clinical Science* **2014,** 126, 305-313.
20. Lubbe, L.; Sewell, B. T.; Sturrock, E. D. The influence of angiotensin converting enzyme mutations on the kinetics and dynamics of N-domain selective inhibition. *The FEBS journal* **2016,** 283, 3941-3961.
21. Waterman, D. G.; Winter, G.; Gildea, R. J.; Parkhurst, J. M.; Brewster, A. S.; Sauter, N. K.; Evans, G. Diffraction-geometry refinement in the DIALS framework. *Acta Crystallographica Section D: Structural Biology* **2016,** 72, 558-575.
22. Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* **2013,** 69, 1204-1214.
23. Collaborative Computational Project 4. The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography* **1994,** 50, 760.
24. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *Journal of applied crystallography* **2007,** 40, 658-674.
25. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* **1997,** 53, 240-255.
26. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **2010,** 66, 213-221.
27. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **2004,** 60, 2126-2132.
28. Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **2010,** 66, 12-21.
29. McNicholas, S.; Potterton, E.; Wilson, K.; Noble, M. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography* **2011,** 67, 386-394.
30. Stanton, J. L.; Watthey, J. W.; Desai, M. N.; Finn, B. M.; Babiarz, J. E.; Tomaselli, H. C. Angiotensin converting enzyme inhibitors: structure-activity profile of 1-benzazepin-2-one derivatives. *Journal of medicinal chemistry* **1985,** 28, 1603-1606.
31. Ksander, G. M.; Erion, M.; Yuan, A. M.; Diefenbacher, C. G.; El-Chehabi, L.; Cote, D.; Levens, N. Dual angiotensin converting enzyme/thromboxane synthase inhibitors. *Journal of medicinal chemistry* **1994,** 37, 1823-1832.
32. Gruenfeld, N.; Stanton, J. L.; Yuan, A. M.; Ebetino, F. H.; Browne, L. J.; Gude, C.; Huebner, C. F. Angiotensin converting enzyme inhibitors: 1-glutarylindoline-2-carboxylic acid derivatives. *Journal of medicinal chemistry* **1983,** 26, 1277-1282.
33. Ksander, G. M.; Yuan, A. M.; Diefenbacher, C. G.; Stanton, J. L. Angiotensin converting enzyme inhibitors: N-substituted D-glutamic acid. gamma.-dipeptides. *Journal of medicinal chemistry* **1985,** 28, 1606-1611.
34. Aoyagi, Y. An angiotensin-I converting enzyme inhibitor from buckwheat (Fagopyrum esculentum Moench) flour. *Phytochemistry* **2006,** 67, 618-621.

Figure Legends

Figure 1: A schematic outline of the ACE binding site occupied by a typical ACEi drug. ACE inhibitors bind to the catalytic region of the active sites of both domains via a chelation interaction with the zinc atom and a hydrogen bond with the catalytic Glu residue. A salt-bridge and hydrogen bonding interactions with the respective Gln, Lys and Tyr residues in the S_2' subsite hold in place the terminal P_2' carboxylic which is attached to a cyclic group. Hydrophobic side-chain mimicking P_1 and P_1' groups interact with the respective hydrophobic S_1' and S_2' subsites. Different regions of the active site are colour coded according to polarity as indicated in the legend.

Figure 2: An illustration of the positional (**A**), metal chelation (**B**) and H-bonding (**C**) docking constraints created and implemented under various permutations when docking ligands against both the N- and C-domains of ACE. Positional constraints (**A**) were set holding either a carbon or nitrogen atom in one of the 1 Å constraint spheres. The metal chelation constraint was created between the zinc atom and a carboxylic acid ZBG from the ligand. The H-bonding constraints were created between both the Lys489/511 and Gln259/281 donors from N-/C-domain and a H-bond acceptor from the P_2' group on the ligand.

Figure 3: Poses of **1** (A and B), **4** (C and D) and **10** (E and F) docked into the N- (cyan; left) and C-domains (green; right). The two key ACEi binding features; a carboxylic acid Zn chelation and a second carboxylic acid acting as an H-bond acceptor in the S_2' subsite are evident. Compound **1** in the N-domain has a twisted peptide bond for the P_1 proline and interactions between the carboxylic acid P_2 group with Tyr369 and Arg381. While in the C-domain **1** contains a trans proline and the P_2 carboxylic acid has no interactions. The bicylic link in compound **4** is too long and rigid to allow for a conventional orientation in either domain, and causes the P_2' and ZBG carboxylic acid groups to bind in different positions in the N- and C-domains respectively. This, along with the flexible alkyl chain allows for a large variation in the position of the P_2 group. Compound **10** is too short to form interactions from its P_2 carboxylic group with Arg381 or Tyr369 in the N-domain, instead it coordinates with the zinc ion. In the C-domain there are no interactions with this this group.

Figure 4: Representative samples of ligand poses from the extended set of diprolyl compounds. A and B, C and D, and E and F show **16**, **17** and **21** respectively docked in the N- (cyan; left) and C-domains (green; right).

Figure 5: A. Schematic representation of compound **16** bound to the N-domain of ACE overlayed with the 2mFo-DFc (blue, contoured at 1σ level) and mFo-DFc (green, contoured at 3σ level) electron density omit maps. The zinc ion is shown as green sphere with the coordinating side chains shown as sticks. **B.** Ligplot representation of the binding site interactions of compound **16** in complex with ACE N-domain. The Arg381 conformation that could possibly provide water mediated interactions with **16** is shown for completeness, although this is not a major form.

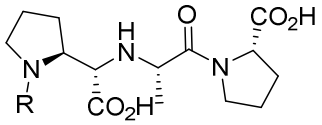
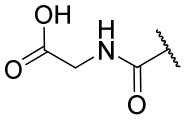
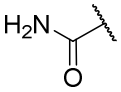
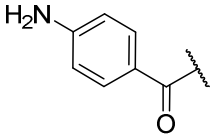
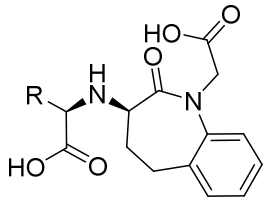
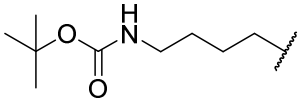
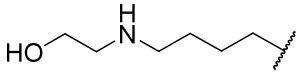
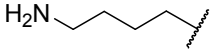
Figure 6. Close up views of the P_2 carboxylic acid binding site. The 2mFo-DFc electron density map (contoured at 1σ level) is shown in blue, the zinc ion and relevant water molecules are shown as green and red spheres respectively. **A-C.** Schematic representation showing the possible alternate conformations of Arg381 in 3 molecules of the complex structure with ACE N-domain. This highlights the weak density for Arg381. **D.** Schematic representation showing the strong interaction between Tyr369 and the P_2 carboxylic acid of compound **16** and the water mediated interaction with Arg500.

Figure 7: A. An overlay of the crystal (orange) and docked (cyan) structures of **16** in the N-domain of ACE. The zinc ions in the crystal and docked structures are shown as green and grey spheres respectively. The two poses of the ligand are closely correlated (RMSD = 1.182 Å) while Arg381 is seen to be highly flexible with two conformations of this flexibility being visible to varying extents in the different chains of the asymmetric units. **B.** An overlay of compound **16**-N-domain crystal

structure (orange) and **16** docked in to the C-domain (green). The zinc ions in the crystal and docked structures are shown as green and grey spheres respectively. The major portion of the scaffolds overlay between the two structures while a shift is observed for P_2 carboxylic acid group showing interactions with Tyr369 and Arg522 for the N- and C-domains respectively. In addition, there is a water mediated interaction with Arg500 in the N-domain.

Figure 8: Overlays of the ACE N-domain complex structures of compound **16** (Orange) with **A** RXP407 (cyan) and **B** 33RE (lilac). The zinc ions are shown as a green sphere for the compound **16** structure and grey spheres for the RXP407 and 33RE structures. Water molecules from the **16** structure involved in mediated interactions are shown as red spheres. The water molecule marked * is conserved in the RXP407 structure in an almost identical position. A water molecule in the 33RE structure that performs a similar mediated interaction is shown as a pink sphere.

Table 1: 13 compounds identified from the GVK database for their polar P_2 groups. These compounds were then docked to assess their potential for selectively inhibiting the N-domain. IC_{50} values given are those reported in the GVK database. The IC_{50} values of compounds **1** and **3** are marked with an asterisk (*) to indicate these values were reported in the database but were later found to correspond to different compounds in the cited reference.

Entry	R	IC_{50} (nM)
	Series 1[18]	
		
1		6*
2		6
3		5.4*
	Series 2[30,31]	
		
4		4.0
5		7.0
6		40
	Series 3[32,33]	

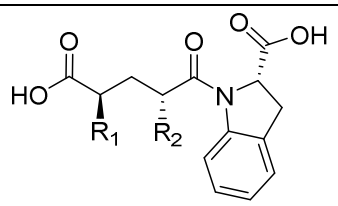
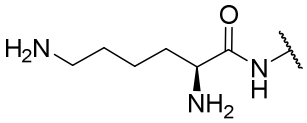
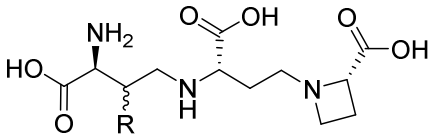
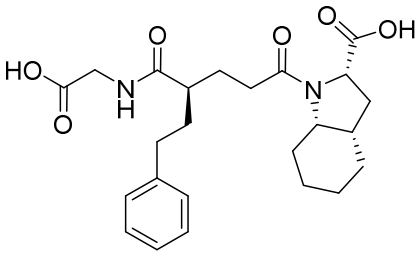
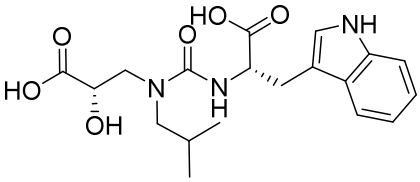
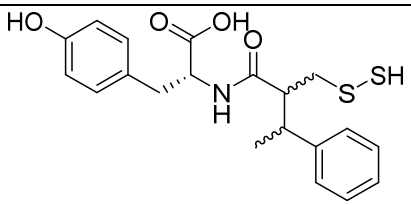
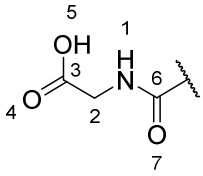
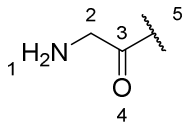
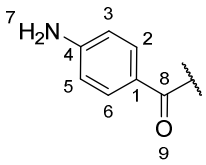
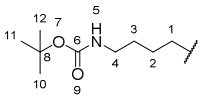
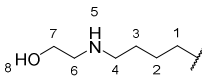
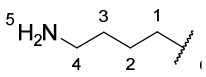
		
7	$R_1 = CH_3$ $R_2 = CH_3$	10
8	 $R_1 =$ $R_2 = H$	188
	Series 4[34]	
		
9	OH	80
10	H	85
	Ungrouped	
11		10
12		103
13		52

Table 2: The three different permutations of the docking constraints employed to help find plausible docking poses of the given ACEi.

Condition	Constraints
C1	Carboxylic acid – Zn chelation, H-bond with Lys489/511 or GlnGln259/281
C2	Carboxylic acid – Zn chelation, amide/amine N in constraint sphere 3
C3	Carboxylic acid – Zn chelation, amide/amine N in constraint sphere 3, amine N or alkyl C in constraint sphere 1

Table 3: Summary of constraint conditions used to dock each compound in the respective catalytic domains. The distances are provided between the nearest ligand atom and an N from Arg381 or an O from Glu403 and the atom from the ligand indicated in brackets.

Entry	N-domain	Distance from Arg381 (Å)	Distance from Tyr369 (Å)	C-domain	Distance from Glu403 (Å)	Distance from Phe391 (Å)
 **1**	C1	2.02 (O-4; Salt-bridge)	2.94 (O-5; H-bond)	C3	5.93 (O-4)	4.22 (O-4)
 **2**	C1	5.62 (N-1)	3.52 (N-1)	C1	7.59 (N-1)	4.45 (N-1)
 **3**	C1	4.05 (N-1)	4.78 (C-2)	C1	5.09 (N-7)	3.95 (C-3)
 **4**	C1	9.15 (C-4)	8.33 (C-4)	C1	6.69 (C-10)	4.74 (C-11)
 **5**	C1	9.74 (C-4)	8.35 (C-7)	C2	11.88 (C-4)	9.77 (C-3)
 **6**	C1	8.19 (N-5)	7.30 (C-2)	-		

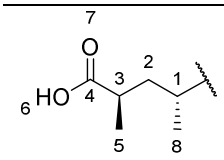
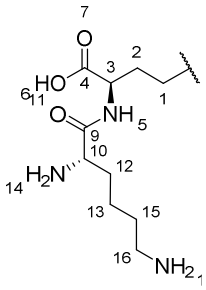
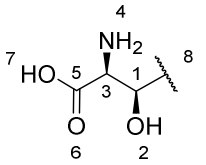
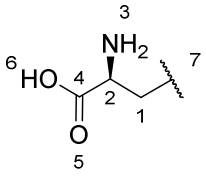
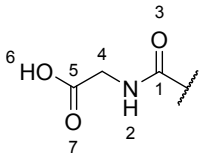
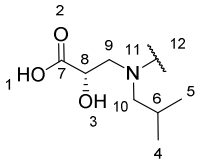
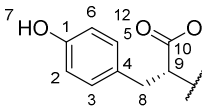
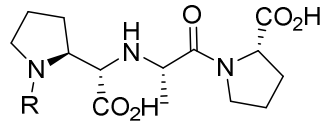
 **7**	C1	9.84 (C-5)	8.13 (C-5)	C1	10.47 (O-6)	9.79 (O-6)
 **8**	C1	8.37 (O-6)	7.53 (O-7)	C2	9.52 (N-14)	6.71 (N-14)
 **9**	C2	8.24 (O-7)	6.73 (O-2)	C2	9.47 (O-6)	8.46 (O-6)
 **10**	C2	5.10 (O-6)	4.16 (O-6)	C2	9.60 (O-6)	8.69 (O-6)
 **11**	-	-		C2	8.82 (C-4)	4.11 (O-6)
 **12**	-	-		C1	9.86 (O-1)	7.93 (O-1)
 **13**	C1	6.70 (O-7)	6.06 (C-6)	C1	9.65 (C-2)	8.43 (C-2)

Table 4: Comparison of the structures found in the GVK database vs the corresponding structures published by Greenlee et al.



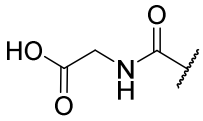
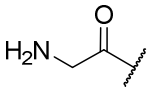
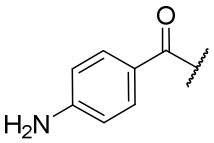
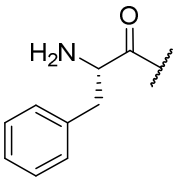
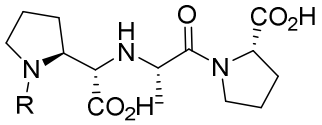
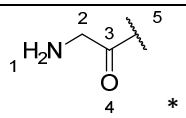
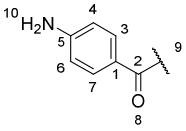
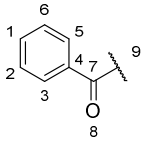
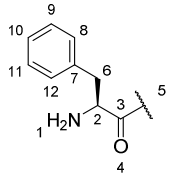
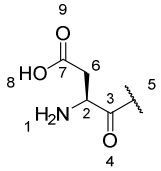
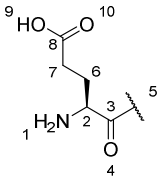
Compound	Incorrect GVK Entry	Compound	Published Structure
1		**2**	
3		**15**	

Table 5: Summary of docking poses, MM-GBSA binding energy calculations and domain selectivity predictions. Each binding pose was assessed by measuring the distance between the ligand and atom indicated in brackets and a N atom from Arg381, an O-atom from Tyr369 in the N-domain with the distances between the nearest aromatic C atom from Phe391 and an O atom of Glu403 from the C-domain. Salt-bridge interactions and raised energy conformations are also noted along with each measurement. The MM-GBSA binding energy is also given as a quantitative metric with which to analyse each pose. Each pose is then judged as non-selective (NS) or N-domain selective (N). Compounds which were selected for synthesis have been marked with an asterix (*).



Compound	Arg381 distance	Tyr369 distance	Glu403 distance	Phe391 distance	C ΔG (kJ/mol)	N ΔG (kJ/mol)	Prediction
 **2***	6.17 (C-2)	5.49 (O-4)	8.36 (O-4)	4.59 (N-1	-52.14	-53.50	NS
 **3***	4.32 (N-10)	2.96 (N-10; H-bond)	5.56 (N-10)	3.28 (N-10)	-62.94	-73.82	N
 **14***	4.12 (C-1)	3.08 (C-1)	6.13 (C-2)	3.90 (C-1)	-47.14	-59.24	NS
 **15***	3.16 (C-11)	3.79 (C-10)	6.27 (C-9)	3.89 (C-9)	-82.72	-56.41	C
 **16***	2.77 (O-9; salt Bridge)	2.95 (O-8; H-bond)	5.56 (O-9)	4.46 (N-1)	-51.57	-71.76	N
 **17***	2.61 (O-10; Salt Bridge)	2.71 (O-9; H-bond)	5.63 (O-10)	3.70 (O-10)	-52.10	-70.32	N

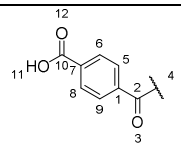
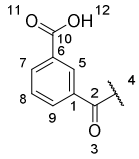
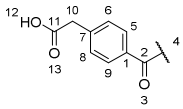
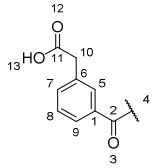
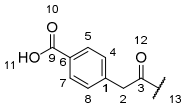
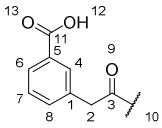
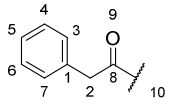
 **18***	2.40 (O-12; Salt Bridge)	2.61 (O-11; H-bond)	5.56 (O-11)	6.25 (C-1)	-54.82	-51.62	N
 **19**	2.60 (O-12; Salt bridge)	2.85 (O-12; H-bond)	6.43 (C-8)	5.28 (C-8)	-43.86	-62.44	N
 **20**	2.63 (O-13; salt bridge)	2.56 (O-12; H-bond)	5.09 (O-12)	6.36 (C-6)	-39.55	-65.84	N
 **21**	2.63 (O-12; Salt Bridge)	2.61 (O-12; H-bond)	4.64 (O-12)	4.08 (C-7)	-48.6	-70.40	N
 **22**	3.03 (O-10; Salt Bridge)	3.48 (C-5)	6.33 (C-5)	3.65 (C-5)	-53.82	-60.10	N
 **23**	2.74 (O-12; salt bridge)	2.59 (O-12; H-bond)	7.41 (C-8)	4.85 (C-8)	-34.90	-58.15	N
 **24**	5.67 (C-2)	5.52 (C-3)	5.20 (C-5)	3.82 (C-6)	-53.46	-53.88	NS

Table 6: Summary of Ki values for each compound synthesised and the observed N-domain selectivity factor. N-domain selectivity is calculated as the C-domain Ki / N-domain Ki. All KI values are reported with a 95% confidence interval.

Compound	N-Domain K_i (nM)	C-domain K_i (nM)	N-Domain Selectivity Factor
2	279.48 ± 46.93	44.38 ± 3.58	0.16
3	3.37 ± 0.55	0.22 ± 0.02	0.07
14	4.45 ± 1.15	0.31 ± 0.11	0.07
15	70.35 ± 9.73	167.80 ± 31.33	2.39
16	11.45 ± 1.37	961.02 ± 153.00	83.93
17	9.63 ± 1.56	17.39 ± 2.38	1.81
18	16.86 ± 1.24	2.29 ± 0.27	0.14

Table 7

	A	**B**
Resolution (Å)	[102.11-9.59] (1.80-1.75)	[101.89-9.86] (1.83-1.80)
Space group	P1	
Cell dimensions (a,b,c) angles (α,β,γ)	74.2, 103.5, 115.5 Å 84.9, 85.5, 82.0°	74.5, 103.4, 115.7 Å 84.5, 85.5, 81.6°
Molecules/asymmetric unit	4	4
Total / Unique reflections	2,276,898 330,946	2,039,183 305,646
Completeness (%)	[99.3] 97.2 (95.7)	[99.1] 97.3 (95.9)
R_{merge}	[0.025] 0.110 (1.366)	[0.028] 0.135 (1.345)
R_{pim}	[0.014] 0.069 (0.933)	[0.016] 0.085 (1.006)
$\langle I/\sigma(I)\rangle$	[29.8] 7.7 (1.0)	[27.4] 7.2 (1.0)
$CC_{1/2}$	[0.992] 0.997 (0.521)	[1.000] 0.997 (0.956)
Multiplicity	[7.8] 6.9 (6.1)	[7.6] 6.7 (5.5)
Refinement statistics		
R_{work}/R_{free}	0.173/0.198	0.193/0.226
Rmsd in bond lengths (Å)	0.009	0.013
Rmsd in bond angles (°)	0.948	1,161
Ramachandran statistics (%)		
Favoured	98.6	98.3
Allowed	1.4	1.7
Outliers	0	0
Average B- factors ($Å^2$)		
Protein	32.0	29.8
Ligand	42.6	41.2
Water	36.3	32.9
Number of atoms		
Protein	39446	39476
Ligand	1930	1897
Water	2410	2381
PDB code	**6EN5**	**6EN6**

Figure 1

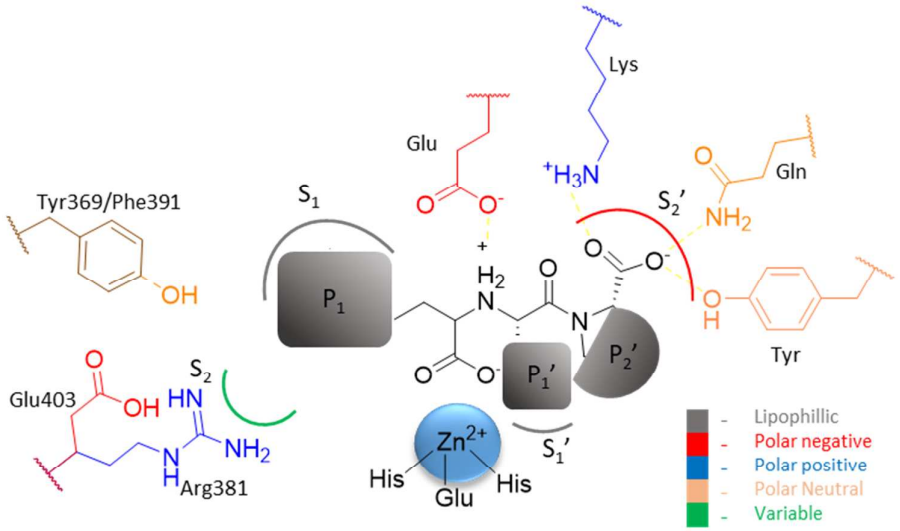


Figure 2

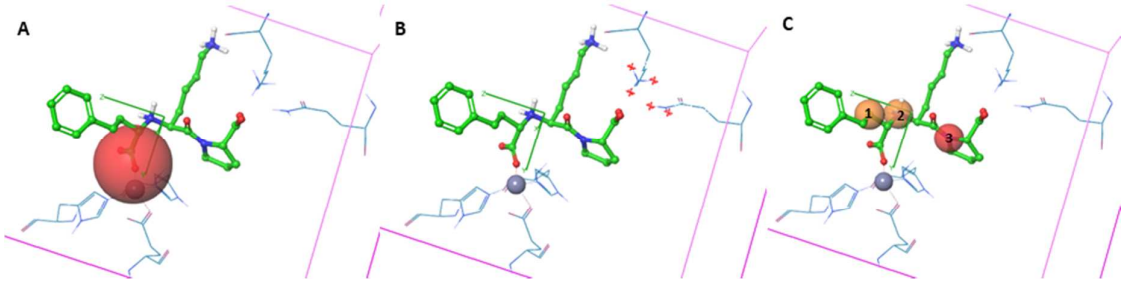


Figure 3

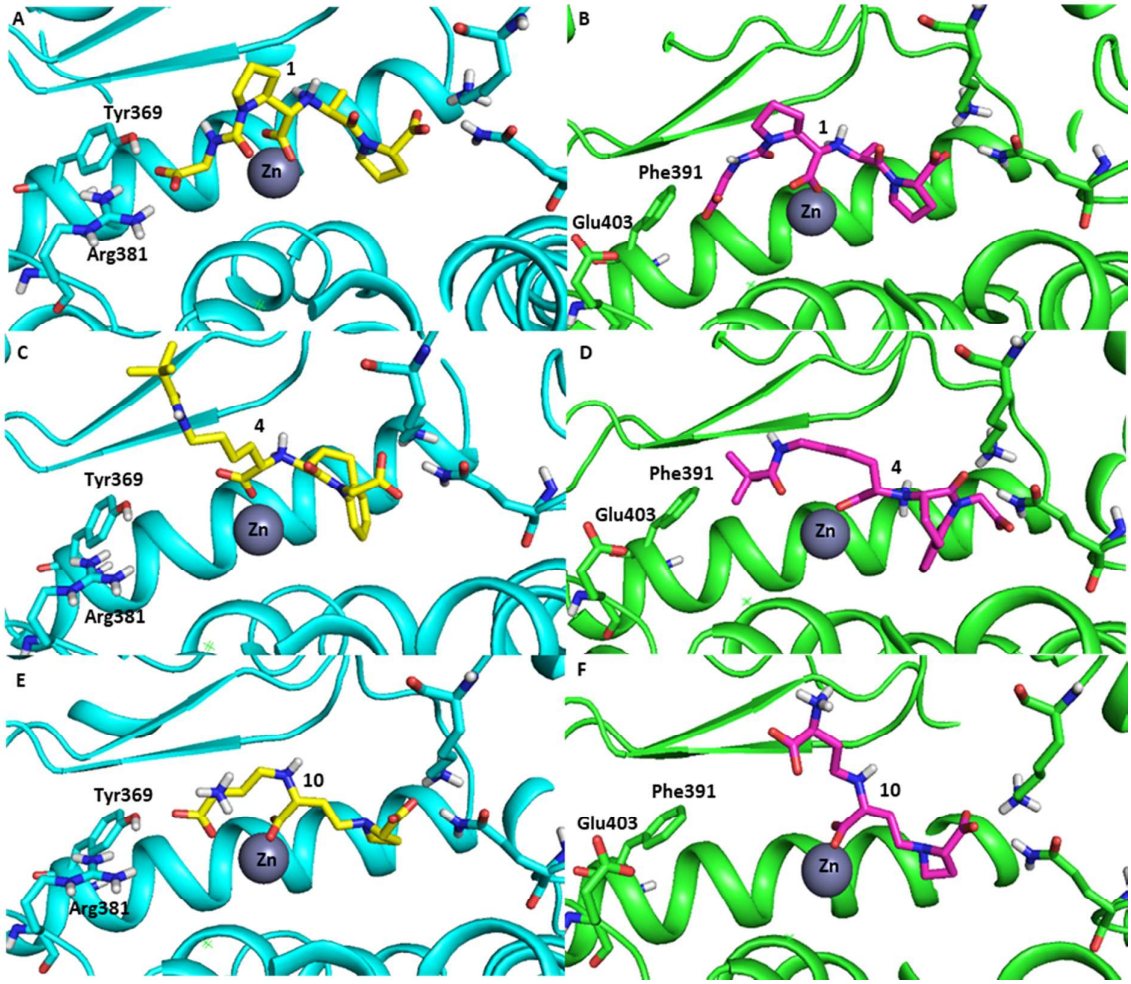

A
1
Tyr369
Zn
Arg381
B
1
Phe391
Glu403
Zn
C
4
Tyr369
Zn
Arg381
D
Phe391
4
Glu403
Zn
E
Tyr369
10
Zn
Arg381
F
Phe391
10
Glu403
Zn

Figure 4

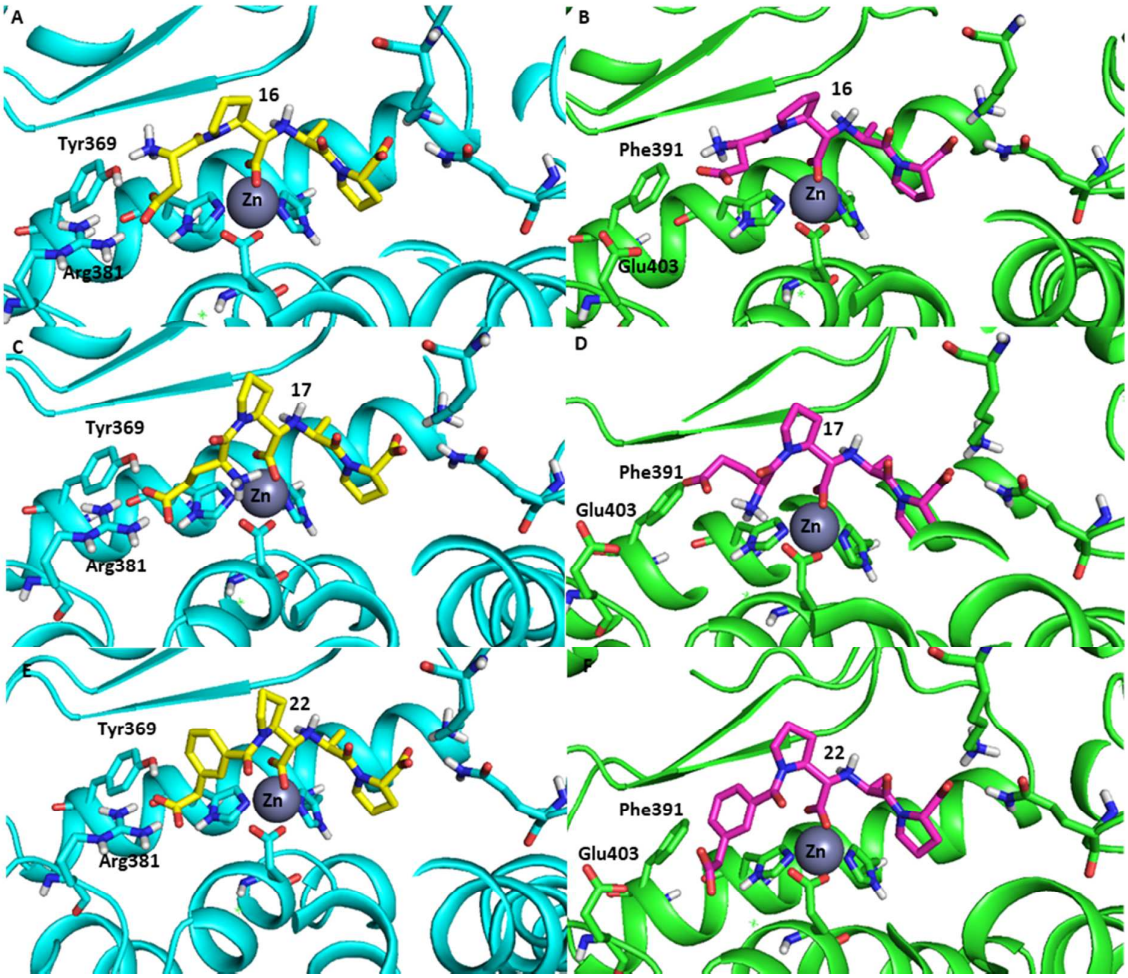

A
16
Tyr369
Zn
Arg381
B
16
Phe391
Zn
Glu403
C
17
Tyr369
Zn
Arg381
D
17
Phe391
Glu403
Zn
E
22
Tyr369
Zn
Arg381
F
22
Phe391
Glu403
Zn

Figure 5

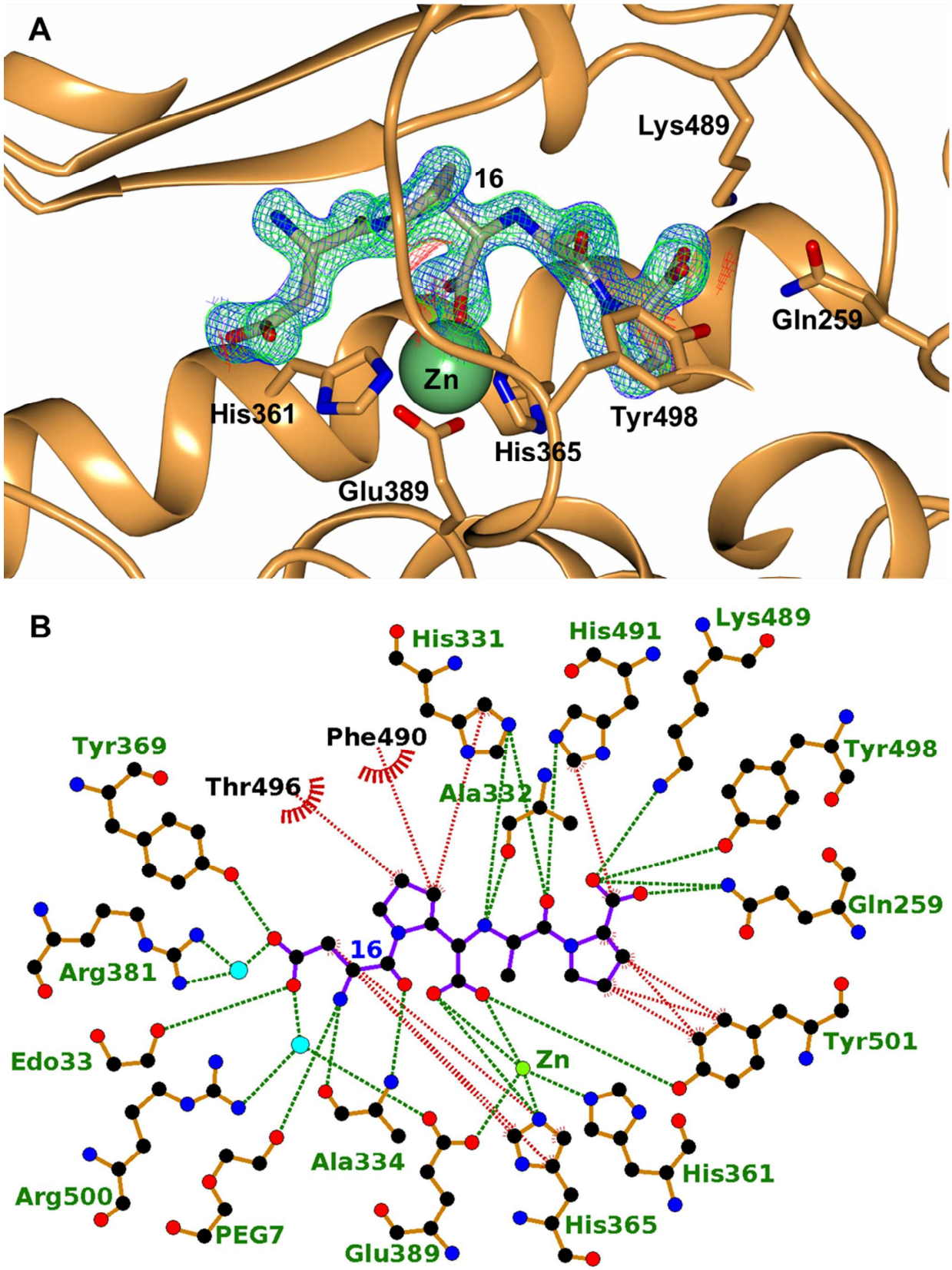

A
Lys489
16
Gln259
Zn
His361
Tyr498
His365
Glu389
B
His331
His491
Lys489
Tyr369
Phe490
Tyr498
Thr496
Ala332
Gln259
16
Arg381
Tyr501
Edo33
Zn
Ala334
His361
Arg500
His365
PEG7
Glu389

Figure 6

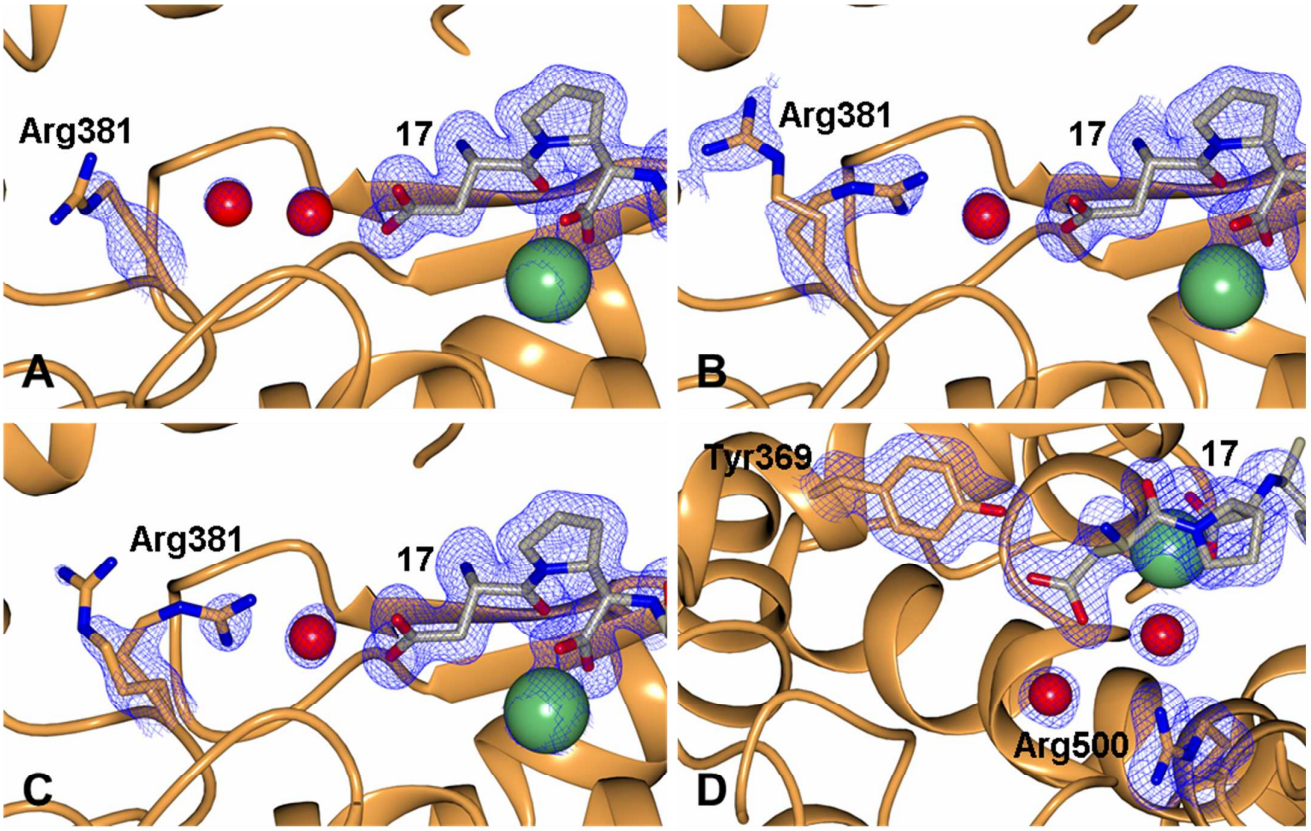


Figure 7

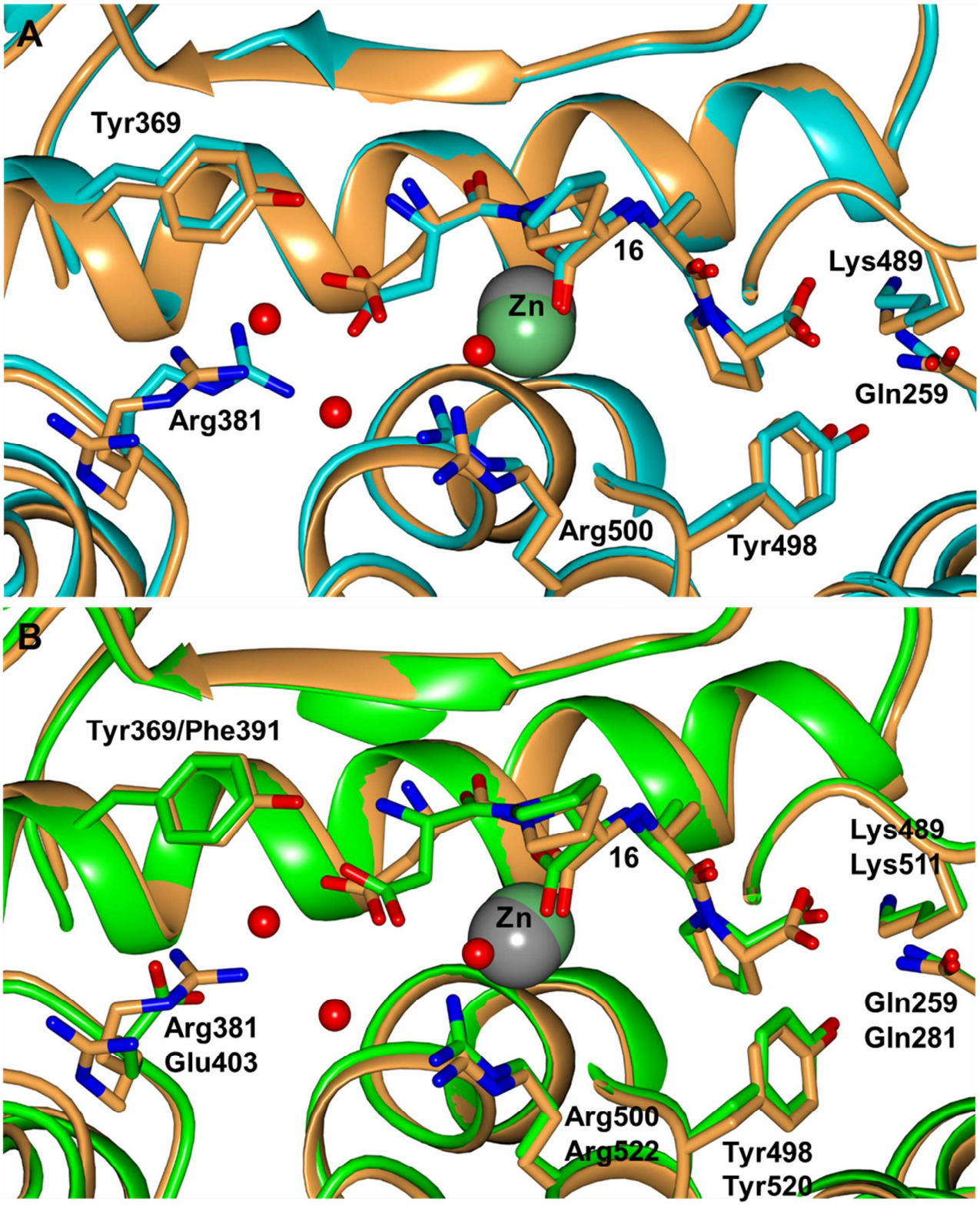
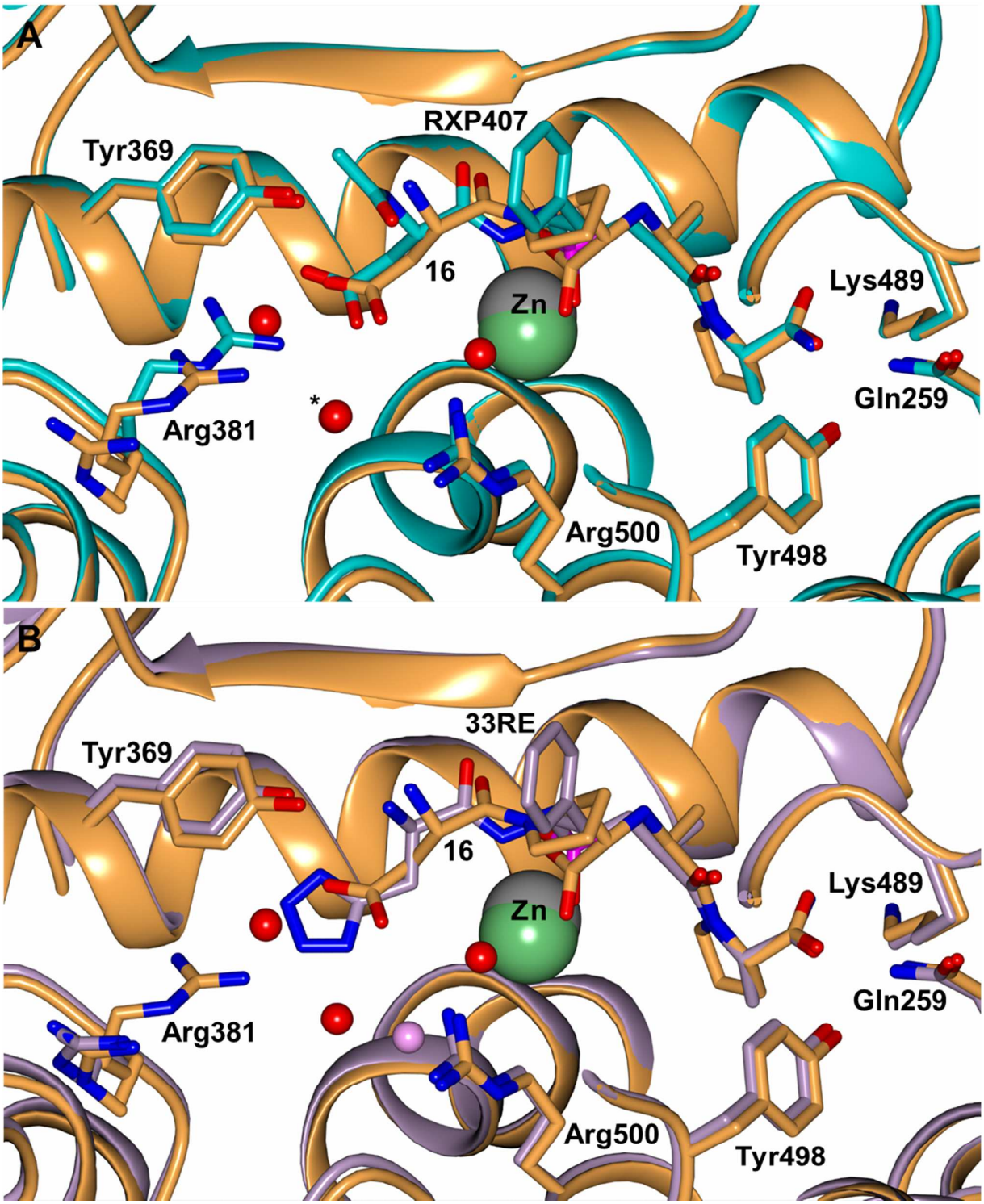

A
Tyr369
16
Lys489
Zn
Gln259
Arg381
Arg500
Tyr498
B
Tyr369/Phe391
16
Lys489
Lys511
Zn
Arg381
Glu403
Gln259
Gln281
Arg500
Arg522
Tyr498
Tyr520

Figure 8


A
RXP407
Tyr369
16
Zn
Lys489
Gln259
*
Arg381
Arg500
Tyr498
B
33RE
Tyr369
16
Zn
Lys489
Gln259
Arg381
Arg500
Tyr498

Scheme 1 – Synthesis of the diprolyl series: i. TMSCN, MeOH/NH4+Cl-, 25 °C 36h, 36% ii. a. HCl in MeOH, 25 °C, 2d. b. $(Boc)_2O$, Et_3N/MeOH, 72% iii. a. Amberlyst 15, MeOH, 60°C, 10 days b. HCl/Dioxane 25 °C, 12hr, 40% iv. a. RCO_2H, T3P, Et_3N, DCM, 0 °C – 25 °C b. LiOH, THF/H_2O, (c. HCl/Dioxane)